Powered by Plain Concepts

#STechDay2021

SINGULARITY

TECH DAY_2021

The era of AI and Cognitive Services

What a “mesh” with my data: from principles to implementation

plain concepts

ORGANIZATION

databricks

SPONSORS

THANK YOU!

Blanca Mayayo

Product Owner Sidra Data Platform at Plain Concepts

With broad experience in IT in different sectors, I am Product Manager for Sidra Data Platform at Plain Concepts, leading the product foundations, go-to-market, and the scaling strategy for engineering.

Previously to my current position, I have transitioned through several engineering and product leadership positions at big companies like Telefónica, Nestlé, and adidas. I consider myself a life-long learner with unavoidable curiosity.

in/blancamayayo/

Index of Contents

Key Trends

Some trends are being consolidated

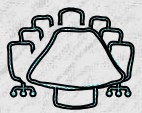
Value and differentiation through data

Product thinking with data

Data as a service

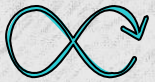
Democratization of AI

Data marketplaces



Data governance & sovereignty

Edge, IoT, etc.



DevOps & automation

Data mesh definition

- How to move from a monolithic Data Lake to a Distributed Data Mesh
  https://martinfowler.com/articles/data-monolith-to-mesh.html

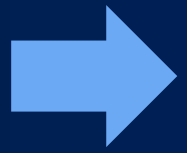
- Data Mesh principles and logical architecture
  https://martinfowler.com/articles/data-mesh-principles.html

Photo by Ave Calvar Martinez, Pexels

What problems does it expect to solve?

symptoms

- No clear ownership or accountability on the data: IT pipelines as “owners” of data
- Lower quality, poorer service metrics
- Difficulty in scaling up the org
- Bottlenecks and friction due to coupled monolithic system

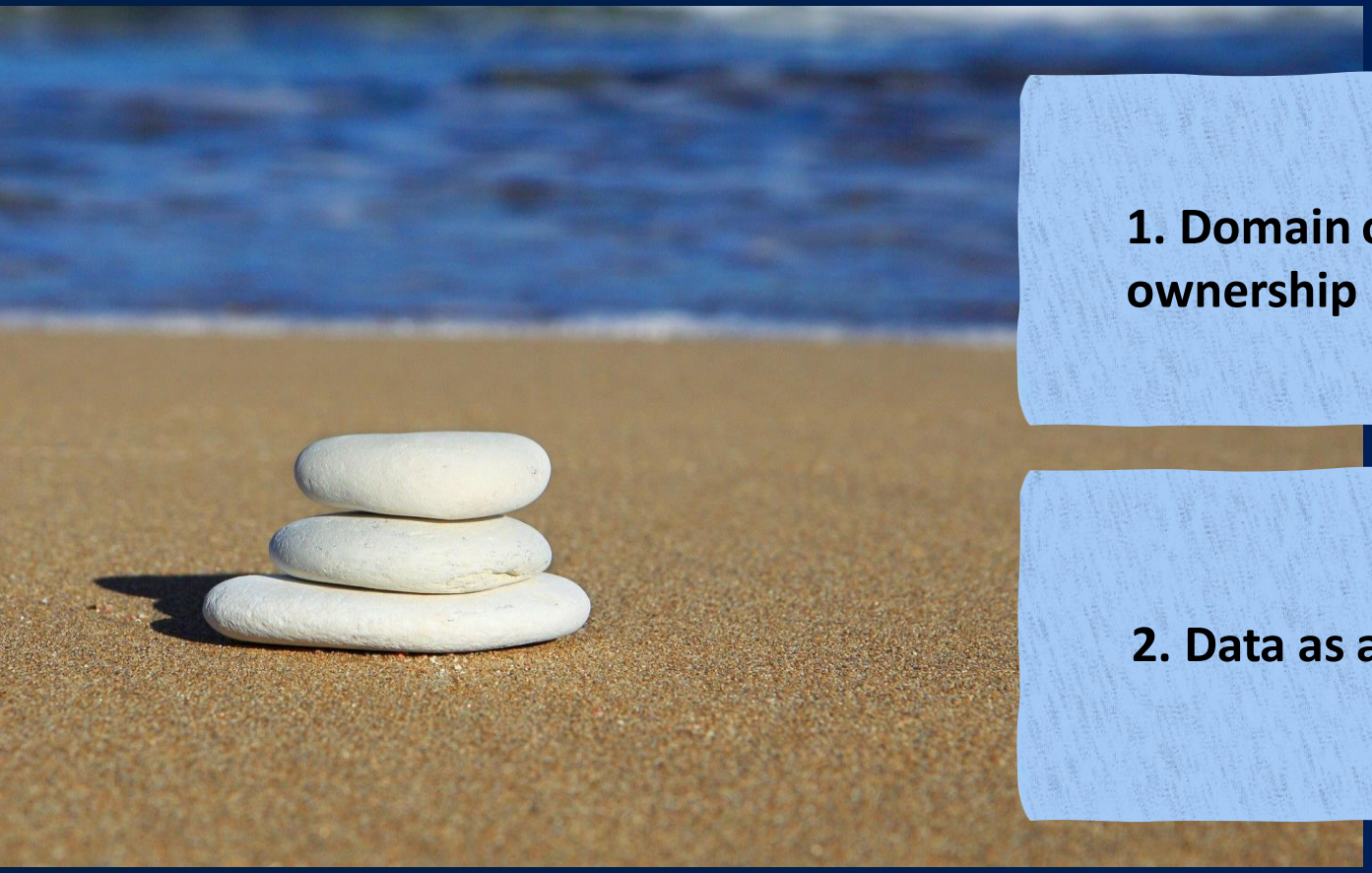


problems

- Siloed data and teams
- Re-work
- Missing feedback loops
- Low usage of the data
- Lack of org. decision / flexibility
- Slow innovation/ value

Key Principles

Key principles to the rescue



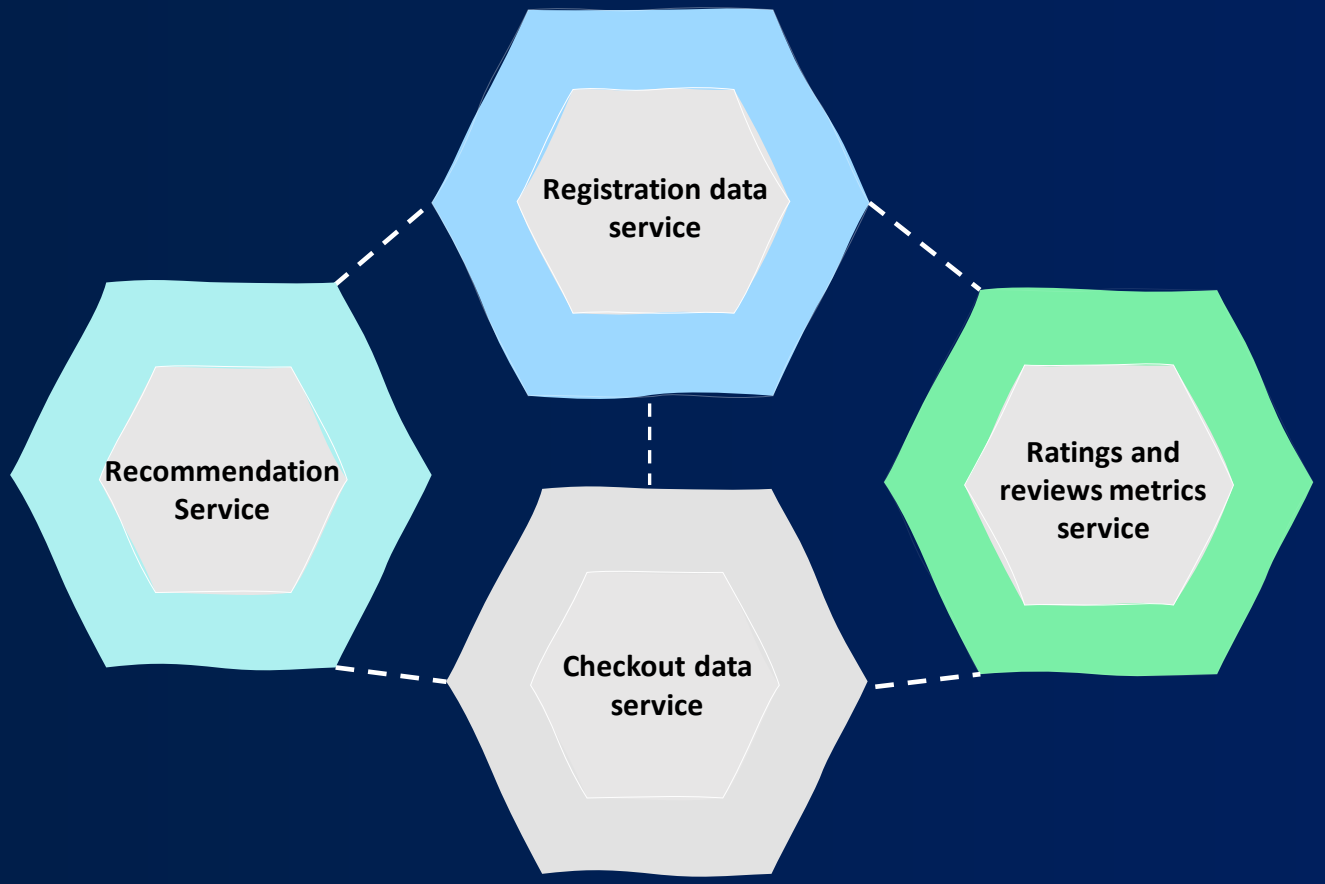
1. Domain oriented ownership

3. Self-serve infrastructure as a platform

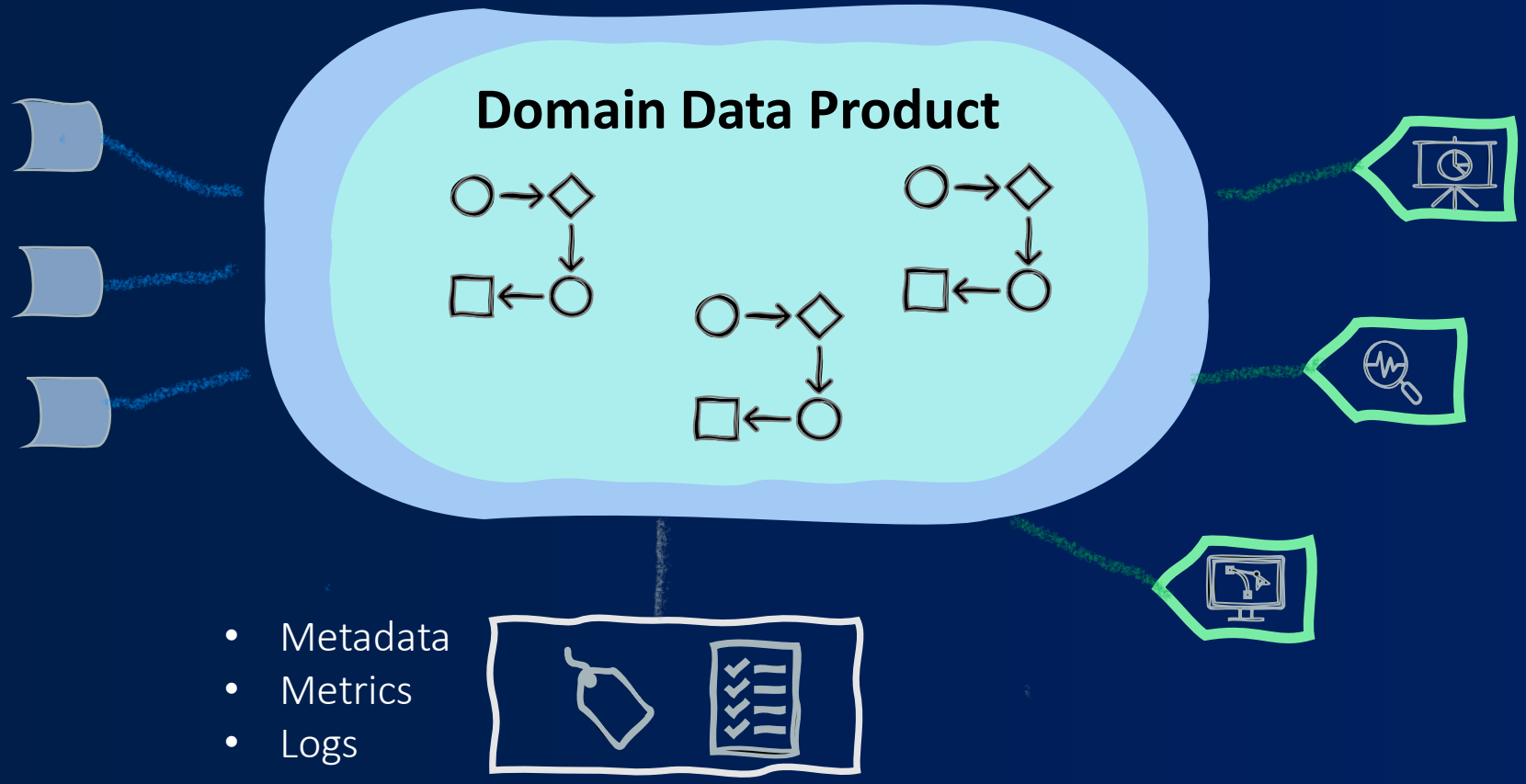
2. Data as a product

4. Federated governance

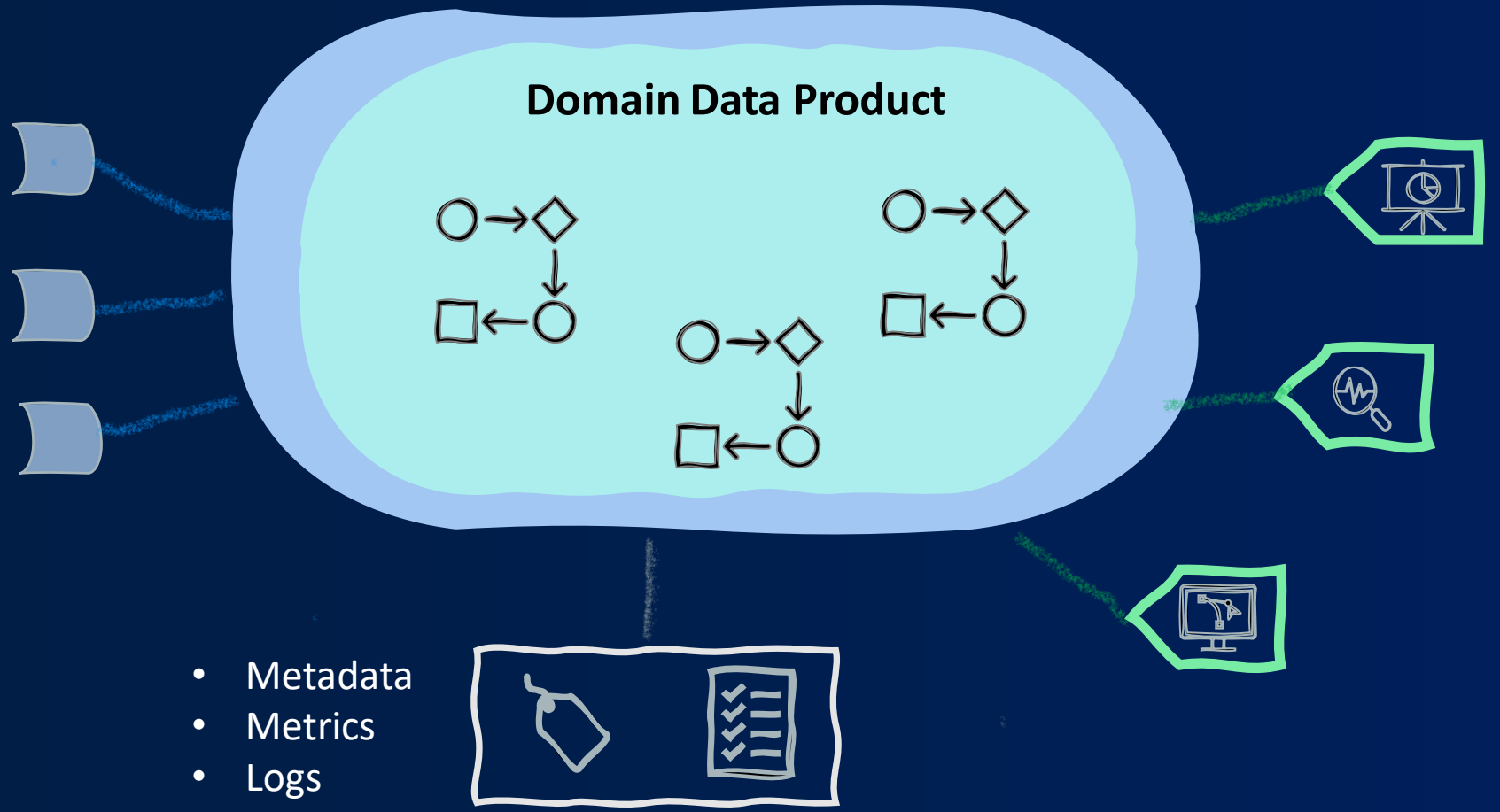
Principle 1: Domain-oriented ownership



Principle 2: Data as a product



Principle 2: Data as a product



- Metadata
- Metrics
- Logs

DATSIS principles

Discoverable

Addressable

Trustworthy

Secure

Interoperable

Self-describing

Data as a product

Customer focus

Deliver business value

Multi-directional feedback

Continuous innovation

Product thinking

Measure success

Continuous improvement

Continuous delivery

One team

Principle 3: Self-serve shared infrastructure

- Domain agnostic, common DataOps tooling
- Centralized governance, but used independently by each domain
- Scaffolding and a lot of automation in place
- Strike the right balance: tight vs loose coupled

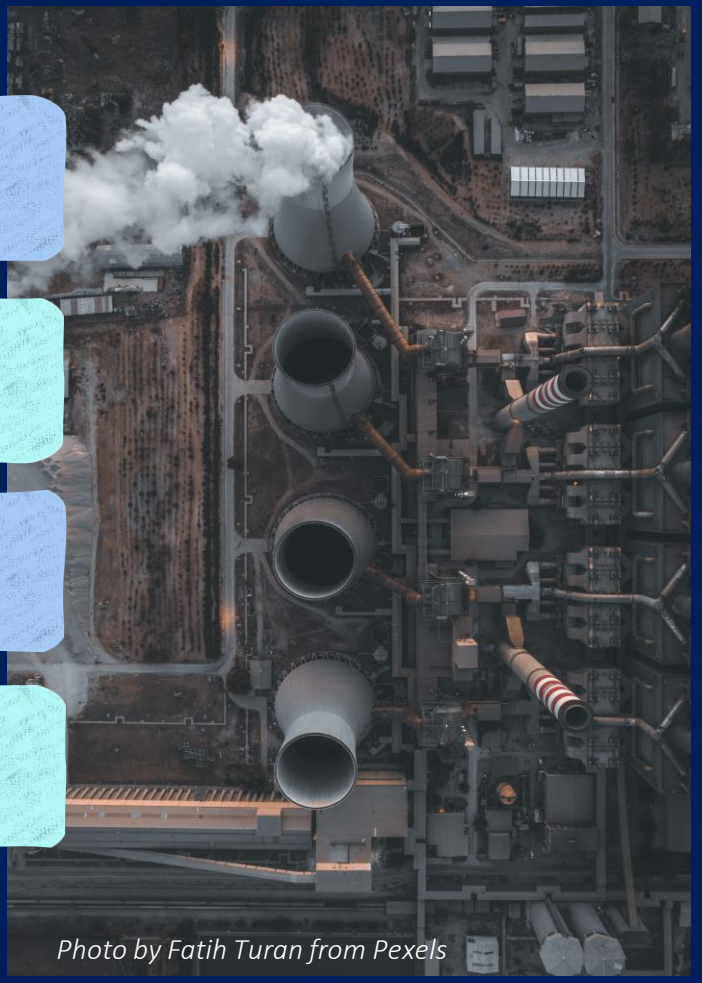
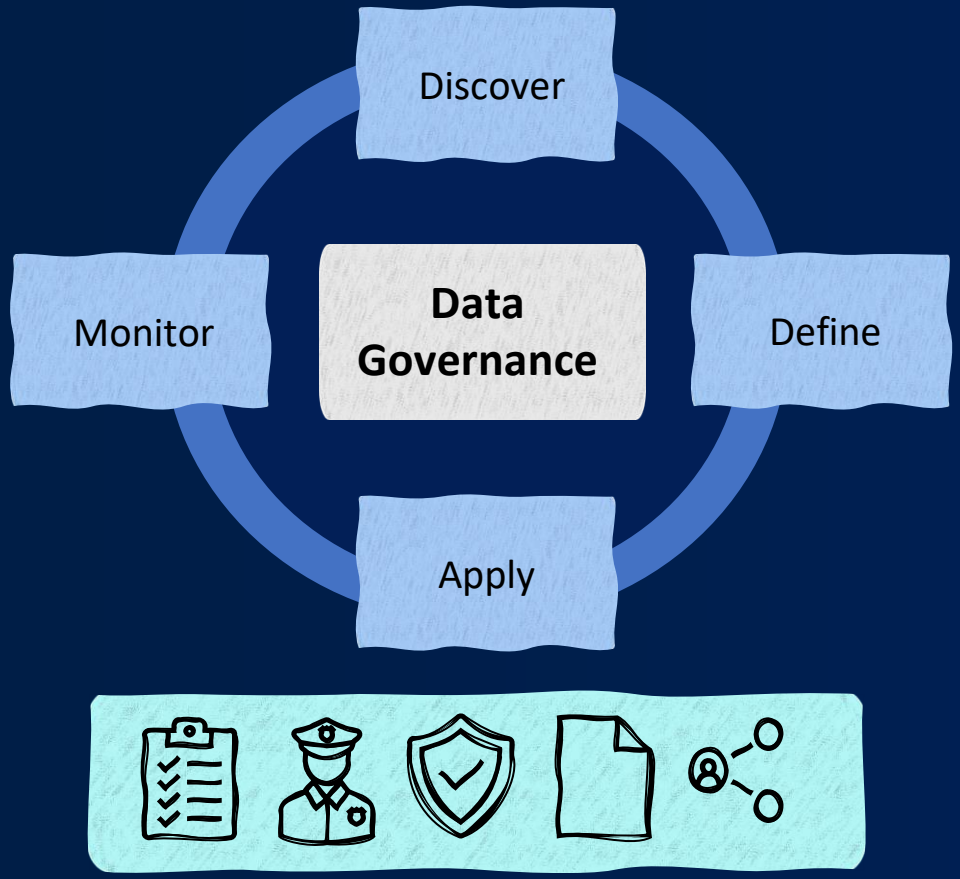


Photo by Fatih Turan from Pexels

Platform Thinking

Principle 4: Federated governance



- Embraces decentralization and interoperability
- Global standardizations across domains
- Centralized data governance vs domain-specific
- Rules for data product versions and compatibility, discoverability
- Backed by metadata and organizational support

Key Considerations

Data mesh is not only technology

- Not technologically prescriptive
- Change management & incentives
- Principles, but leave the implementation open (evolutionary)
- Processes, DevOps, DataOps best practices
- Knowledge and skills management

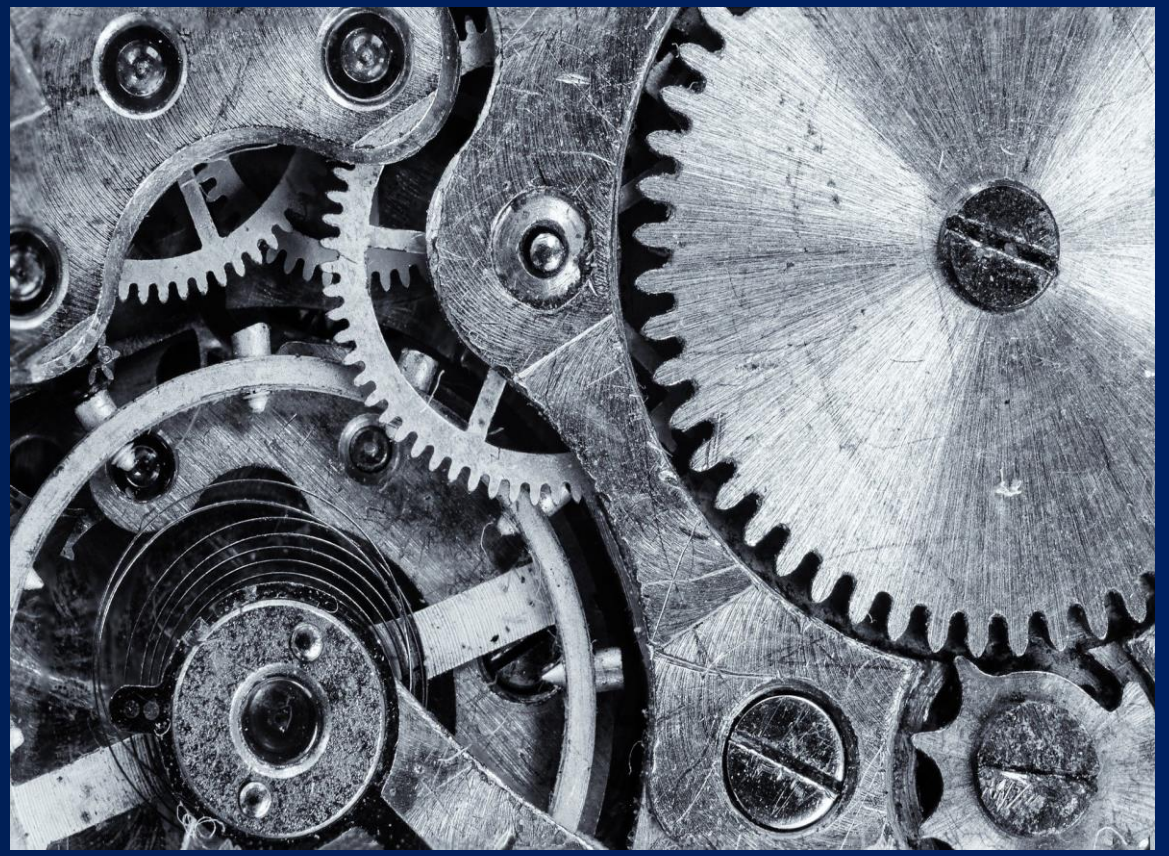


Photo by Pavlofox from Pixabay

Data mesh is not for everyone

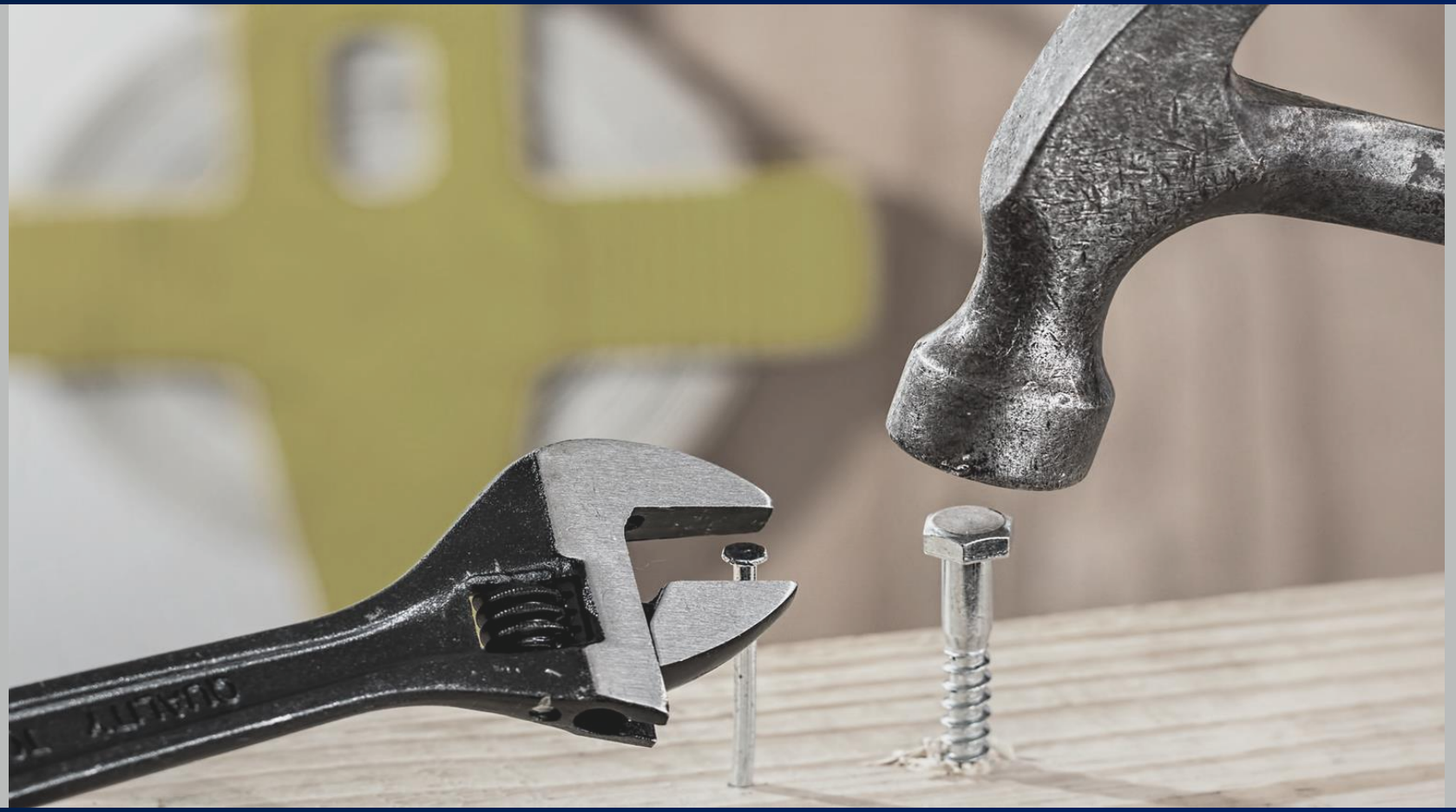


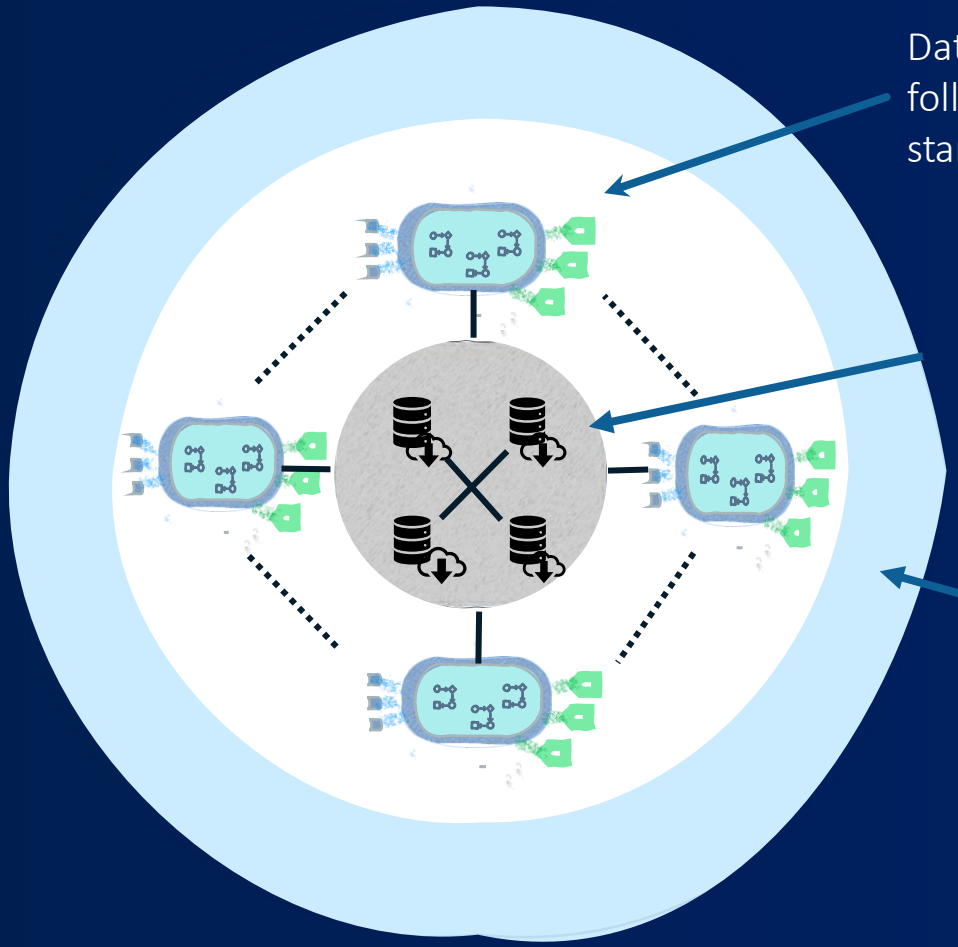
Photo by Steve Buissinne from Pixabay

There is not ONE data mesh

Foto by Olga Lioncat in Pexels

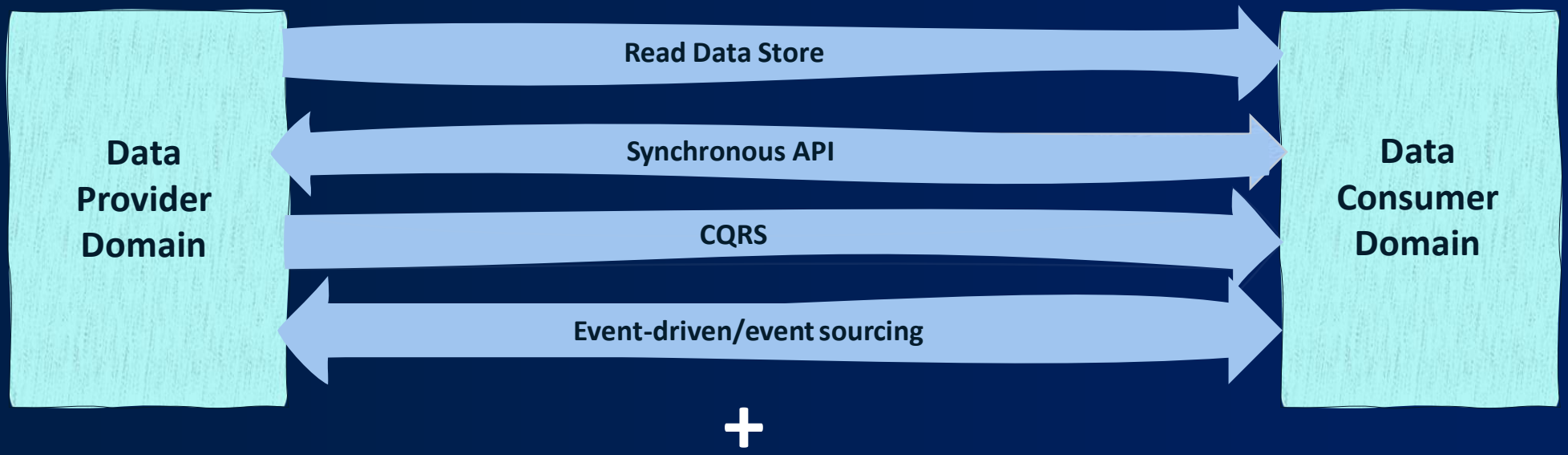
There is not ONE data mesh

Different degrees of centralization and decentralization for raw and transformed data



Towards implementation

Beyond data distribution or application to an application distribution framework



Support for automated deployments and updates, granular access control, elasticity, extensibility, observability, discoverability,…

A common toolbox

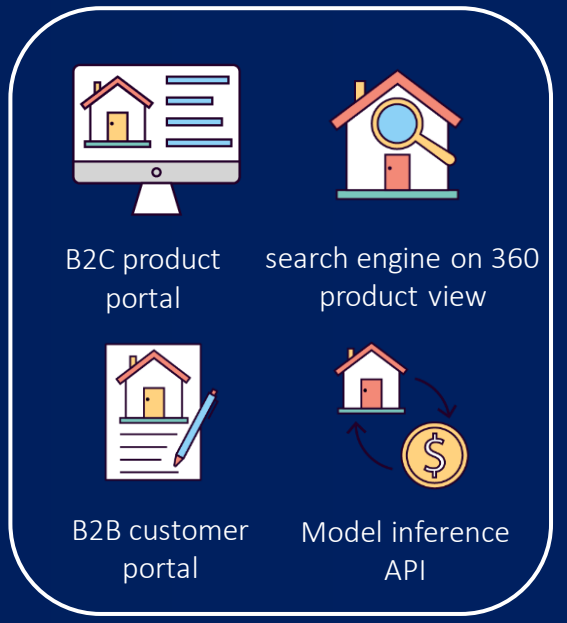
- **Data discovery:**
  - Search, metadata catalogue service, lineage
- Scalable **storage and compute**
- **Data governance:**
  - Security, access control management, federated identity management
- **Data build**:
  - Common blueprints/scaffolding for use cases for data transformations, model training
- **Data operations:**
  - Pipeline orchestration as code, infra as code, logging and monitoring, DevOps, Use cases deployment, Model management and deployment

A practical use case

Our use case in the real estate sector

Source systems, operational plane

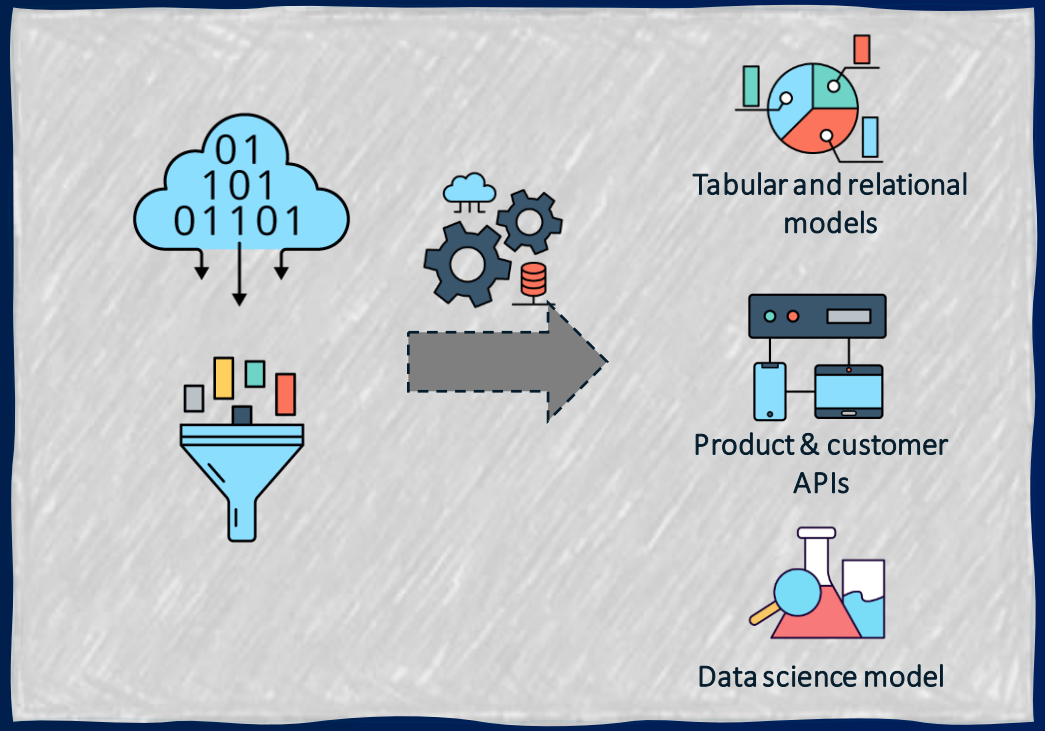
Achieve a **mesh of use cases** that better resemble the **business model** and value provided to B2C and B2B customers



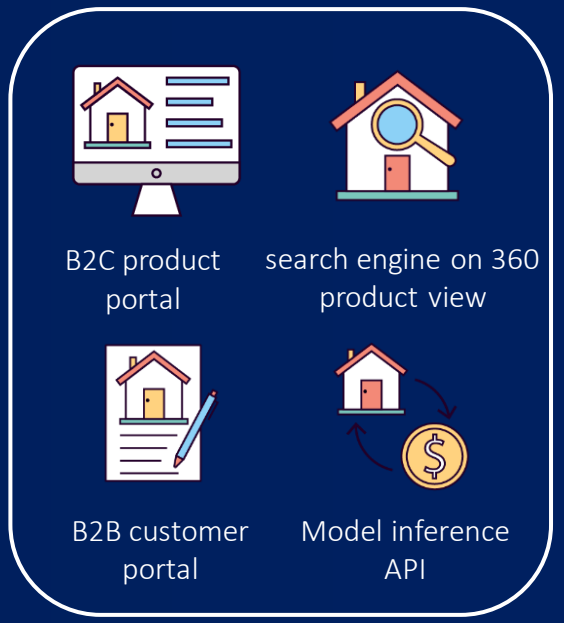
End-user products, business plane

Our use case

Source systems, operational plane



Multi-interface data layer of Interoperable data products



End-user products, business plane

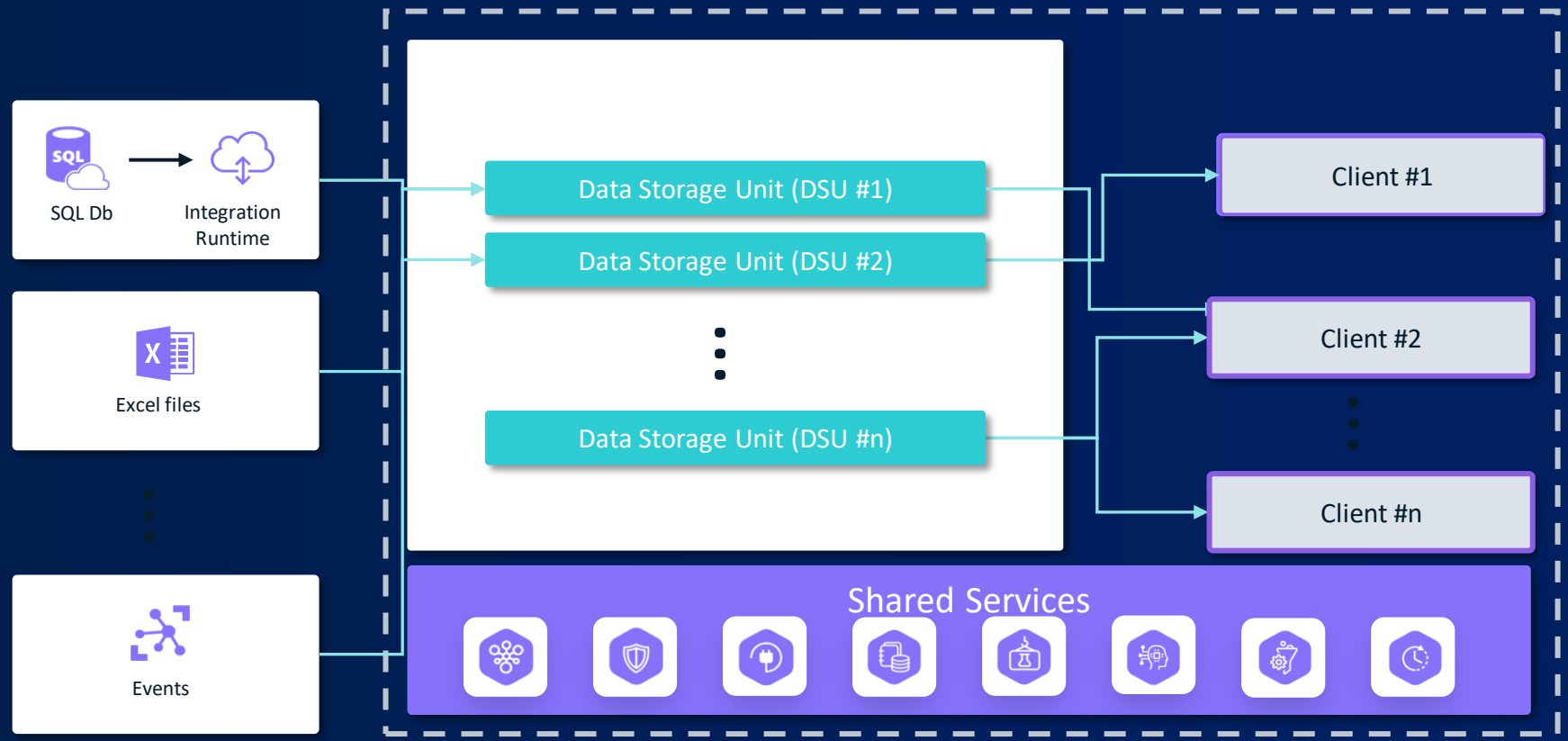
Our implementation

Our building blocks: data mesh runway

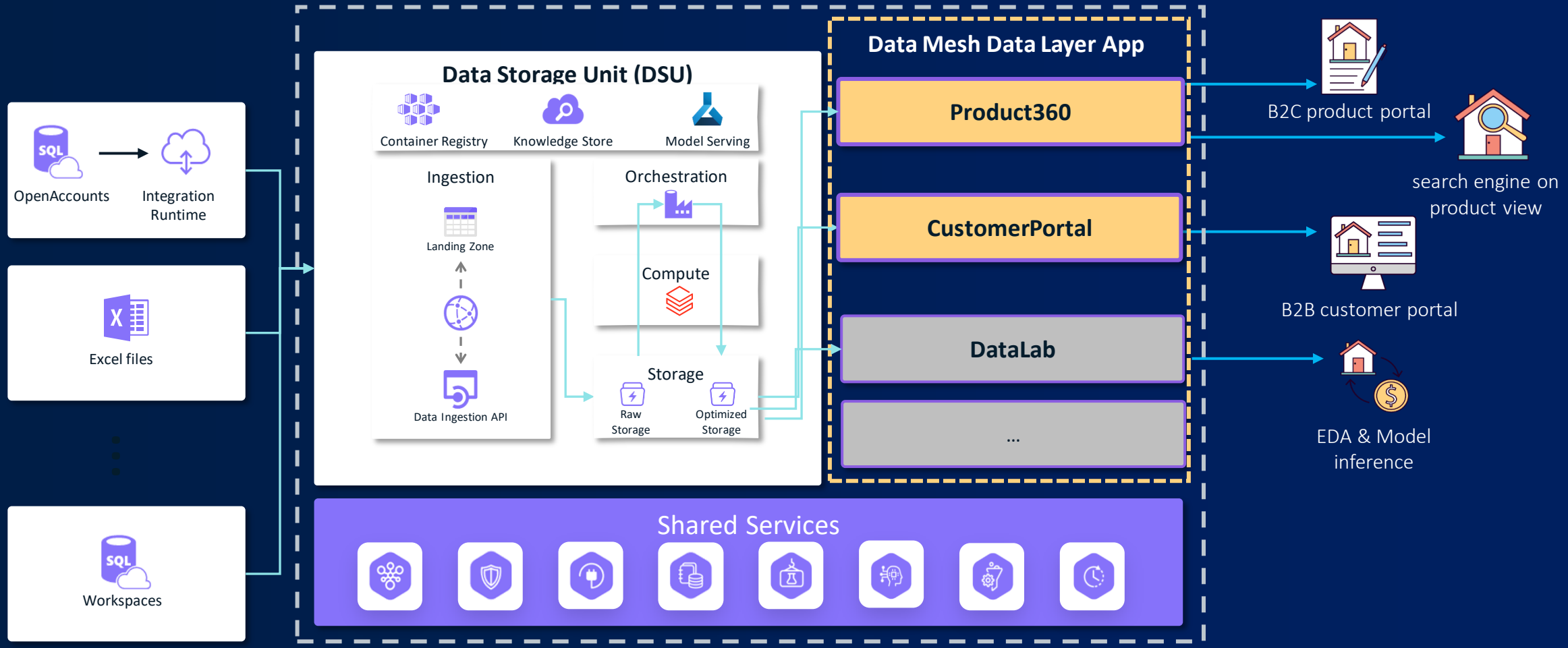
Data-mesh compatible data platform

Raw data in optimized format, with granular access control

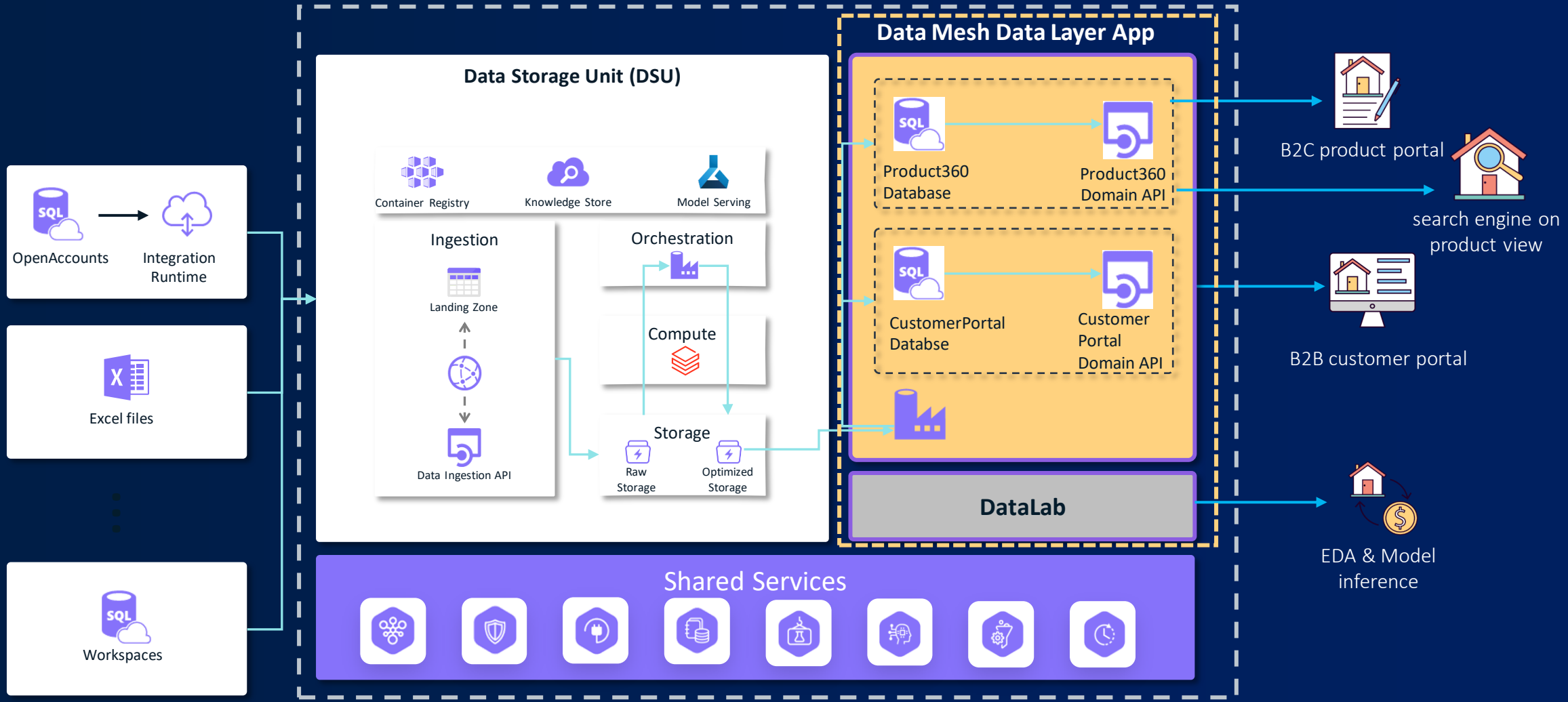
Each business data product is a domain in a Client Application.



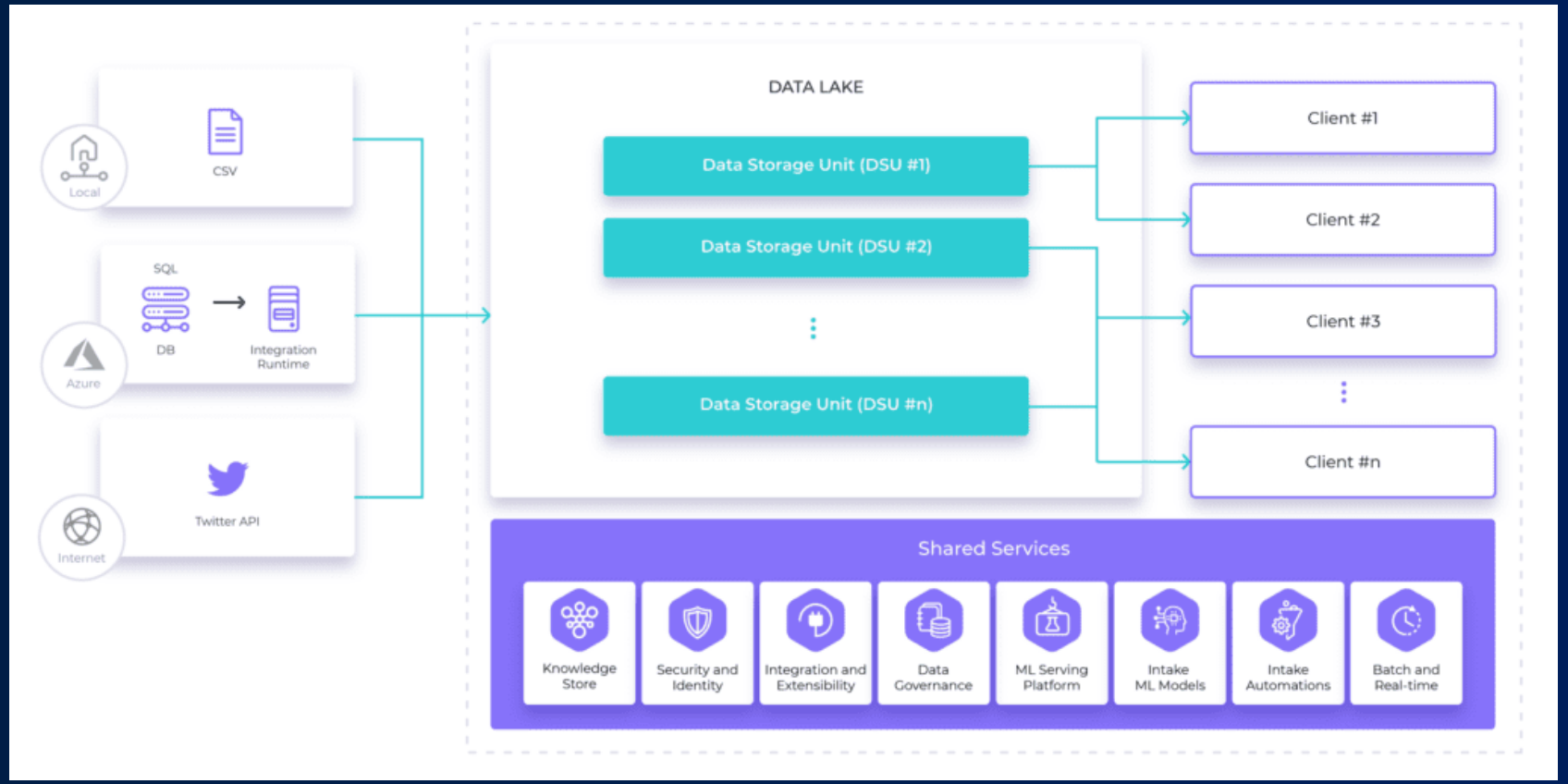
Our building blocks: Client Applications



Our building blocks



Our business case accelerators



Interoperable Client Apps as **accelerators of governed use cases** thanks to a **shared services and infrastructure:**

- Data intake accelerators (e.g., automated pipelines, infra as code)
- Common security model
- Management APIs and Web UI
- Metadata model and API
- Data catalogue
- Common monitoring and metrics
- DevOps enablers (CI/CD, provisioning, etc.)
- Client Application blueprints with embedded services

Key Takeaways

- **Cross-functional domain teams** and **product thinking** with data
- Data **management and governance** (glue to interoperate).
  - Policies, tooling, automation
- **Self-serve** compatible platform (platform thinking) to lower the barrier of usage and empower teams

SINGULARITY
TECH DAY_2021

Thank you!